DOCUMENT RESUME

ED 109 227 TM 004 688

AUTHOR	Samuels, S. Jay; Edwall, Glenace E.
TITLE	Measuring Reading Achievement: A Case for Criterion-Referenced Testing and Accoutability. NCME Measurement in Education, Vol. 6, No. 2, Spring 1975.
INSTITUTION	Michigan State Univ., East Lansing. Office of Evaluation Services.
PUB DATE	75
NOTE	8p.
AVAILABLE FROM	Office of Evaluation Services, Michigan State University, East Lansing, Michigan 48823 (Subscription, $2.00 per year; Single Copy, $0.50, 25 or more, $0.35 ea.)
EDRS PRICE	MF-$0.76 PLUS POSTAGE. HC Not Available from EDRS.
DESCRIPTORS	Achievement Tests; *Criterion Referenced Tests; Decision Making; *Educational Accountability; *Educational Assessment; Norm Referenced Tests; Reading Achievement; *Reading Instruction; Reading Tests; Standardized Tests; Student Evaluation; Testing; *Testing Problems

ABSTRACT

The accountability movement in American education has received great clamors of attention in the past few years. This movement, in turn, suggests the need for particular data to inform responsible decision-making; the need for assessment instruments which can address the question of what is learned; and the need for strengthening the existing ecosystem between schools, universities, the public, and the government agencies concerned with education. Specifically, problems within the context of reading instruction and assessment are discussed. Assessing students in relation to a criterion of mastery; pointing out existing weaknesses in specifying objectives, selecting and designing evaluation instruments, interpreting evaluation data and improving instructional methodology are advocated strongly. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

"PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL BY MICROFICHE ONLY HAS BEEN GRANTED BY
NCME
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED109227

TM 004 688

Measuring Reading Achievement: A Case for Criterion-Referenced Testing and Accountability*

S. Jay Samuels

Glenace E. Edwall

ABOUT THIS REPORT

The accountability movement in American education has received great clamors of attention in the past few years. Dr. S. Jay Samuels, Professor of Educational Psychology, and Glenace E. Edwall at the University of Minnesota discuss the problems within the context of reading instruction and assessment. They strongly advocate assessing students in relation to a criterion of mastery; point out existing weaknesses in specifying objectives, selecting and designing evaluation instruments, interpreting evaluation data and improving instructional methodology. Positive suggestions are offered for strengthening the existing ecosystem between our schools, universities, the public and the government agencies concerned with education.

Dr. Samuels is Director of the Minnesota Reading Research Projects and has done extensive research in reading acquisition.

Glenace Edwall is a Fellow in the Center for Research in Human Learning, is involved in research in the reading project, and consults on program evaluation locally.

CJF

S. Jay Samuels and Glenace E. Edwall
University of Minnesota

Evaluation in American education has entered a new era in recent years with the clamor of public interest groups for information about their schools, and the recognition by taxpayers and educators alike that the school is in some sense accountable for the products of the system. The Minneapolis Citizens League, typifing the community movements for educational accountability, notes

> *We have come to understand that there are two primary clients to be served by our public schools: society, and the individual student (and/or the family as spokesman or guardian of the student). . . . Both society at large and the individual student's family have a very legitimate claim on public education. . . . The prevalent form or level of accountability is no longer accepted as being sufficient. Increasingly, the constraints of limited resources coupled with the desire for excellence has generated a demand for a results-oriented system of accountability. Put simply, people want to know what outcomes are achieved by the expenditure of educational resources. At the same time, many parents and students are dissatisfied with the degree to which their school system is responsive to them.*[1]

As this call implies, the type of "accountability" that has traditionally been operative, i.e., answering specific charges of parents when difficulties arise and informing the public only of "input information" such as teacher preparation and budget allocations for equipment and buildings,[2] is not sufficient for dealing

*Support for this paper was provided by grants to the Center for Research in Human Learning (NSF, NICHD, The University Graduate School), the Minnesota Reading Research Project (NICHD) and the Research & Development Center for the Educationally Handicapped (BEH).

with the central question of the accountability movement, which is, "What is learned?" To approach answering that question, we would suggest, demands that the schools have the tools for making responsible decisions in response to the needs of society and individual students, and these tools include the specification of objectives, means of assessment commensurate with those objectives, and the development of available channels for effecting changes in and upgrading the quality of instruction. As we shall argue, currently these tools are largely not employed to the best advantage: at all levels, objectives are less than clear; the tests most frequently used are not directed to answering the what is learned question, and thus cannot provide information for decision-making; and the ties between the school districts and the resources of university colleges of education are tenuous. Changes in relation to these needs must be made in order that the schools can be accountable to the public which is their support and to the students who are their responsibility.

"Our testing is largely misused and non-functional."

The Problem of Assessment

It is the problem of assessment which is central to the demand for accountability, for it is only to the extent that educators can measure learning (and lack of such) by their students that they can report to the public. Although no country in the world currently employs standardized achievement testing to the degree of the U.S., paradoxically we still do not have answers to basic questions of what students have learned. Our testing is largely misused and non-functional.

The misuse of standardized tests has become commonplace, fanning the anti-test movement which has resulted in court decisions barring the administration of familiar achievement/intelligence tests in some areas. Farr and Roser outline the most frequent misuses of such standardized tests as including test administration without a clearly stated purpose; use of tests related to specific goals for the assessment of global objectives; use of test results as the sole criterion for judging an educational program; improper release and interpretation of results; and use of tests as classifying tools for a rigidly-tracked, labeled system of educational stratification. Farr and Roser conclude,

> *Taken together, these five misuses of tests and test results are justifiably significant evidence for universal opposition to the continual 'misuse of tests' in our nation's schools.*[3]

Despite the justifiable outrage against misuse of tests, however, means of assessment continue to be demanded. Thus the more serious charge against the tests is that even with appropriate safeguards in administration and interpretation, the tests most widely used may have no relation to the objectives of instruction and are therefore non-functional (and, thus, in the extreme case, may actually be severely damaging to an individual student's educational progress and society's right to know the outcomes of education).

". . . norm-referenced tests . . . are not of value . . . for responsible decision-making."

The essence of the problem of assessment instruments such as standardized reading achievement tests lies in norm-referencing. The most commonly used reading achievement tests, sample items of which we shall examine in some detail, are norm-referenced tests; the data that can be provided from such tests is of a comparative nature only and cannot answer the what is learned question or provide direction for decision-makers.

The philosophy and beginning use of norm-referenced tests came from an era when classification of individuals was the raison d'etre of testing; early intelligence tests, as is well known, were designed for the purpose of separating those who could benefit from the contemporary educational system from those who could not, and testing found an early function and expansion in making similar classification decisions for industry and the military. It is immediately obvious that the purpose of these tests and the form of the scores they reported was for determining who was better than whom at whatever the test ostensibly measured.

When such tests are taken into the classroom the charges which can be leveled against the practice are legion. Because of the emphasis on placement of one person in relation to others, it is the total score (or total scores for subtests) which is of primary importance, and the tests thus provide no specific knowledge of students' competencies, no diagnostic information in terms of specific difficulties, and no information which can guide decision-making relative to the school's responsibility to the individual student.

More specifically, the emphasis on total score frequently results in the mixing of items which are generally related to the labeled score but may be based on rather different abilities, e.g. literal and inferential comprehension items in reading achievement tests. Not wishing to single out any one publisher's test for the kind of generic problem commonly found in numerous commercial reading tests, we offer the following hypothetical example:

Lost in the Woods

John and Bill carefully slid their boat onto the muddy land. They jumped ashore followed by

their dog. For hours the boys and their dog wandered through the woods looking for the beaver pond. As the sun started to set John and Bill became aware they were lost. Bill called the dog and told him to go back to the boat. The dog sniffed at the trail as he ran through the woods and in a short time led them back to the boat.

Comprehension Questions:

1. *In the woods, the dog followed the trail by*
   a. *sight*
   b. *smell*
   c. *touch*
   d. *sounds*
2. *The boys were in the woods looking for a*
   a. *lost child*
   b. *lost dog*
   c. *beaver pond*
   d. *lake*
3. *John and Bill were*
   a. *glad their dog was along*
   b. *brothers*
   c. *much too young to be hiking in the woods*
   d. *foolish to cross the lake in a boat*

There are several problems with these comprehension questions. Question one can be answered without having read the paragraph, by means of general knowledge of canine behavior. Question two measures literal comprehension of information contained in the passage. Question three measures inferential comprehension since information about the correct answer is not contained directly in the paragraph and the student taking the test must make an inference about the response alternative which is most probably correct.

If a student fails to answer question three correctly one cannot easily diagnose the nature of the problem. The student may be able to interpret literally what he reads but be poor in inference reasoning. On the other hand, the student's problem may be failure in literal comprehension which would, of course, prevent him from reasoning from the information provided in the paragraph.

As is often the case in norm-referenced tests, a single comprehension score is assigned to the student. This score provides virtually no diagnostic information about the student's reading strengths and weaknesses. The score is primarily useful in comparing the student to others of similar age, but this score is of limited usefulness. At the very least, norm-referenced tests should provide two comprehension scores for each student, one for literal and the other for inferential comprehension.

In that the items which are retained through the development of the norm-referenced test are those which are highest in predictive validity, they may actually be low in content validity: "good" items are those which 50 per cent of the test takers can pass, but the items which are most useful in such a discrimination are not the most highly related to instructional goals (presumably high priority material which was taught well should be passed by 90 per cent of students).

As Ralph Tyler sums this problem:

> *These tests thus provide dependable information about where the child stands in his total test performance in relation to the norm group. But when one seeks to find out whether a student who made a low score has learned certain things during the year, the test does not include enough questions covering the material on which he was working to furnish a dependable answer to that question.*[4]

One may conclude, then, that the norm-referenced tests widely employed to gauge educational progress in reading and other basic skills areas do not approach the "what is learned" criterion for assessment, and thus are not of value to the educator concerned about data for responsible decision-making.

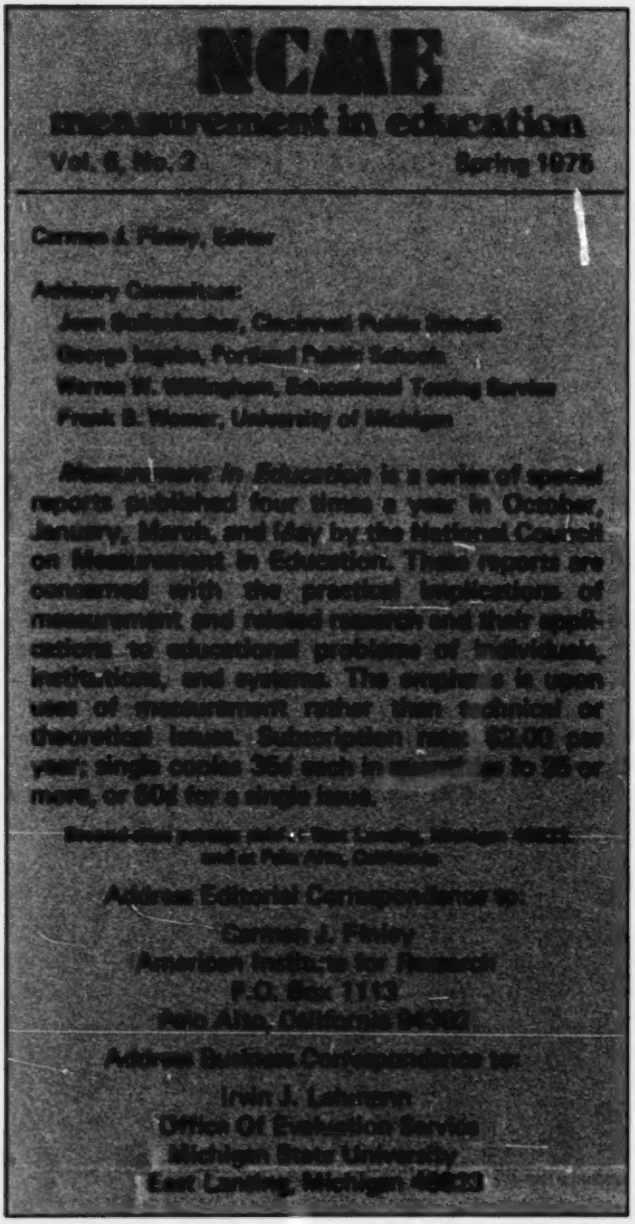
One Solution: The Move to Criterion-Referenced Testing

Viewed in this way, the most obvious solution to the assessment problem is to design instruments sensitive to *what* students have learned rather than their relative achievement. This approach is termed criterion or domain-referenced testing, derived from the emphasis which is placed on the individual student's standing in relation to a criterion of mastery of a given skill or subject matter. Put most simply, the criterion for test item selection in this model is mastery, rather than discrimination; "the test developer is not interested in the spread of performance but rather in how many students are able to perform well enough to pass the anchor point."[5] It is our contention that such design can better serve educational decision-making needs.

". . . a more complex technological society cannot afford the luxury of educational rejects . . ."

The conceptual framework for criterion-referenced testing is based on the mastery model of instruction promulgated by Carroll, Bloom and others.[6] The assumptions inherent in this model are simply that most children can learn the content of instruction, and that the societal needs shaping educational and testing decisions are different than in the era of norm-referencing: in the basic skills areas with which we are primarily concerned, a more complex technological society cannot afford the luxury of educational rejects, and must insure that its children all have a mastery of necessary skills. Thus we assume that children can learn, and our task in testing is measuring each child's progress toward a skill/knowledge criterion.

The support for these assumptions and the move to criterion-referenced testing comes from the models of



NCME

measurement in education

Vol. 6, No. 2 Spring 1975

Carmen J. Finley, Editor

Advisory Committee:

Jean Bolesta-Lukes, Cincinnati Public Schools

George Ingles, Portland Public Schools

Warren W. Willingham, Educational Testing Service

Frank B. Womer, University of Michigan

Measurement in Education is a series of special reports published four times a year in October, January, March, and May by the National Council on Measurement in Education. These reports are concerned with the practical implications of measurement and related research and their applications to educational problems of individuals, institutions, and systems. The emphasis is upon uses of measurement rather than technical or theoretical issues. Subscription rate $2.00 per year; single copies 35¢ each in quantities of 10 to 25 or more, or 60¢ for a single issue.

Second-class postage paid at East Lansing, Michigan 48823 and at Palo Alto, California

Address Editorial Correspondence to:

Carmen J. Finley

American Institutes for Research

P.O. Box 1113

Palo Alto, California 94302

Address Business Correspondence to:

Irvin J. Lehmann

Office Of Evaluation Service

Michigan State University

East Lansing, Michigan 48823

educators such as Carroll and Bloom, and from research work with populations presenting special instructional problems. Conceptually, the model proposed by Carroll

> *. . . makes it clear that if the students are normally distributed with respect to aptitude for some subject . . . and all the students are provided with exactly the same instruction (same in terms of amount of instruction, quality of instruction, and time available for learning), the end result will be a normal distribution on an appropriate measure of achievement. . . . Conversely, if the students are normally distributed with respect to aptitude, but the kind and quality of instruction and the amount of time available for learning are made appropriate to the characteristics and needs of each student, the majority of students may be expected to achieve mastery of the subject.*[7]

". . . ghetto children are learning to read . . . where reading and spelling instruction are emphasized and subskills are learned to mastery"

That this model for mastery is in fact workable can be seen in research and reading programs designed for populations with special instructional problems; in a variety of situations, pupils who would be consigned on the basis of norms to "slow learner" or "low achieving" groups are being taught to read. Samuels and Dahl, reviewing a reading program in the Kansas City schools, concluded that ghetto children are learning to read in schools with a philosophy of success where reading and spelling instruction are emphasized and subskills are learned to mastery;[8] and a variety of community reading programs for special target populations have been meeting success when "each program is designed around the needs of the students."[9]

Recent research has found virtually no difference in performance on very simple learning tasks among individuals who seemingly differ considerably in IQ; Samuels and Anderson found no differences between two IQ groups of third graders in a simple associational learning task,[10] and Zeaman and House found an attentional variable among three groups of retarded learners in a discrimination learning task, but the three acquistion curves showed no differences once the critical features of the stimuli had been perceived.[11] Other studies of basic learning ability have similarly found no difference in actual tests of learning performance among individuals who differed considerably in IQ.[12]

If we are therefore convinced that most children can learn the content of instruction, and that it is desirable both for the student and for society that he do so, the purpose of testing is not assessing students' relative standing, but rather measuring their progress toward mastery of the unit of instruction in terms of a given criterion. Note that with this approach, if instructional objectives and the commensurate form of assessment are clearly defined and well designed, the data from testing are both specifically related to students' competencies, i.e., attempts to answer the what is learned question, and can provide diagnostic information in revealing the student's particular strengths and weaknesses in a skill or knowledge area. This is so because items are specifically constructed with content validity as the most important criterion, as opposed to discriminatory or predictive ability.

Concern with finding out what students know in a more absolute sense has motivated a variety of criterion-referenced approaches to testing and an

emphasis on the content of items and their relation to instructional objectives. The National Assessment of Educational Progress is the most far-reaching of such projects as a plan for a "systematic, census-like survey of knowledge, skills, understandings and attitudes designed to sample four age levels in ten different subject areas,"[13] and the items developed by NAEP exhibit the concern with measuring what students know in relation to the objectives of instruction:

> *The most important criterion which was established for exercise development was that every exercise must be a direct measure of some knowledge, skill or attitude which was stated in the objectives. That is, it must have content validity. . . . An exercise must be meaningful, make sense and be directly related to the objective. It must not be trivial, inconsequential or peripheral to the objective.*[14]

State education agencies are initiating similar assessment programs for measuring pupil achievement through the development of educational objectives and construction of tests to measure how well these objectives are being met in current educational programs,[15] and criterion-referenced tests and guidelines for corresponding objectives are also beginning to be available commercially for classroom use.[16] Thus, although the criterion-referenced approach is not without problems of its own, particularly in the time of development, it is beginning to be usefully employed to provide information on the outputs of our instructional systems.

". . . the accountability movement has drawn attention to the need for stronger ties between the various parties involved in education. . . ."

In addition to the necessity of providing new forms of assessment, the accountability movement has also drawn attention to the need for developing stronger ties between the various parties involved in education: the schools, the universities' colleges of education, concerned citizenry, and relevant governmental agencies.

The present system regarding these entities can well be termed a deteriorating ecological interaction. In the schools, as noted above, testing is often misused and non-functional; test results often are not reported publicly; "research" consists of making compilations from the school's archives rather than either correlational study or experimentation; and, thus, with objectives not stated explicity and no public knowledge of assessment results, the factors necessary for accountability are simply not present.

Problems also exist in this ecosystem between the schools and colleges of education. The colleges of

REPORTS AVAILABLE

Back issues of *Measurement in Education* are available at 35¢ each in quantities of 25 or more for a single issue.

Vol. 1, No. 1	*Helping Teachers Use Tests* by Robert L. Thorndike
No. 2	*Interpreting Achievement Profiles—Uses and Warnings* by Eric F. Gardner
No. 3	*Mastery Learning and Mastery Testing* by Samuel T. Mayo
No. 4	*On Reporting Test Results to Community Groups* by Alden W. Badal & Edwin P. Larsen
Vol. 2, No. 1	*National Assessment Says* by Frank B. Womer
No. 2	*The PLAN System for Individualizing Education* by John C. Flanagan
No. 3	*Measurement Aspects of Performance Contracting* by Richard E. Schutz
No. 4	*The History of Grading Practices* by Louise Witmer Cureton
Vol. 3, No. 1	*Using Your Achievement Test Score Reports* by Edwin Gary Joselyn & Jack C. Merwin
No. 2	*An Item Analysis Service for Teachers* by Willard G. Warrington
No. 3	*On the Reliability of Ratings of Essay Examinations* by William E. Coffman
No. 4	*Criterion-Referenced Testing in the Classroom* by Peter W. Airasian and George F. Madaus
Vol. 4, No. 1	*Goals and Objectives in Planning and Evaluation: A Second Generation* by Victor W. Doherty and Walter E. Hathaway
No. 2	*Career Maturity* by John O. Crites
No. 3	*Assessing Educational Achievement in the Affective Domain* by Ralph W. Tyler
No. 4	*The National Test-Equating Study in Reading (The Anchor Test Study)* by Richard M. Jaeger
Vol. 5, No. 1	*The Tangled Web* by Fred F. Harcleroad
No. 2	*A Moratorium? What Kind?* by William E. Coffman
No. 3	*Evaluators, Educators, and the Publics: A Détente?* by William A. Mehrens
No. 4	*Shall We Get Rid of Grades?* by Robert L. Ebel
Vol. 6, No. 1	*On Evaluating A Project: Some Practical Suggestions* by John W. Wick

education have a resource of experts in subject matter areas of relevance to the schools, but these are chiefly concerned with teacher preparation, graduate training, and research. Where service to school districts does occur, the expert is in the role of an entrepeneur; the system is also marked by few concerted efforts on the part of the schools to bring the educational experts directly to the task of improving the educational attainment of their students. As regards the state agencies concerned with education, few systematic attempts have been made to collect data on educational achievements.

"In the schools . . . the factors necessary for accountability are simply not present."

Some changes, however, are already being made, fostered by the accountability movement, which are strengthening the interaction; we shall note these and add some suggestions:

1. The state and the schools: Some state assessment in reading is now being done, using in part materials developed by the National Assessment of Educational Progress.[1] Such data could also be collected for school systems by the state agency, through financial "piggy-backing" on the state program by local school districts, using funds currently spent for standardized assessment instruments. In this way, data would become available for each school and age group in the school district, and could be used to inform both the public and decision-makers. Such a "piggy-back" program is now being used in Richfield, Minnesota schools, drawing on state assessment materials in reading.

"(There are) few concerted efforts . . . to bring the educational experts directly to the task of improving . . . Educational attainment . . . of students."

---

2. Colleges of education and the schools: Although the type of interaction which would be optimal between these systems is prolonged, systematic, developmental interest in the schools by educational experts, it must be realized that the university staff is simply not sufficiently large for such direct involvement. Therefore, we would suggest that the staff could be best utilized in setting up model programs in troubled school districts. These programs, in turn, may serve as demonstration centers, e.g., for reading instruction, to a wider audience, and would allow for the systematic collection of data on experimental methods and programs. Further, these programs could then be extended by especially training graduate students to implement them in school districts needing help; this practice would also provide future educators with valuable applied experience.

3. The schools and accountability: In order to have the data to become accountable, as noted above, we have suggested that the schools need help in specifying objectives, selecting and designing evaluation instruments, interpreting the data of evaluation, and designing and improving instructional methodology. We have suggested fundamental changes in calling for decision-making based on data collection for diagnostic information; for changes in evaluation and testing including the selection of texts and tests on the basis of the criterion/mastery model and participation in state assessment programs with "piggy-backing" to obtain specific data on district schools; and for aid in instructional design via closer ties with university colleges of education. To these suggestions we would add two further points:

". . . input from educators, subject matter specialists, and concerned citizens could be extremely helpful for school districts . . ."

Central to the problem of accountability (once we can assume that at least some direction has been given to the measurement question) is the need for the school district, with the input of the citizens concerned, to specify the objectives for which accountability is held. A model for developing objectives can be found in California schools under the provisions of the Stull Bill, where a mutually teacher- and principal-prepared plan evaluates teachers on the basis of the actual performance of pupils in achieving the formulated objectives.[18] The methodology of the National Assessment of Educational Progress in developing objectives may also provide a helpful guide for the schools:

". . . there are no incentives . . . for teachers and administrators in the current system."

---

1. *The objectives must be satisfactory goals for each subject area as seen by subject matter specialists.*
2. *The objectives must be ones which currently are accepted as goals of American education by most schools.*

3. *The objectives must be ones which are acceptable to thoughtful lay adults as reasonable goals of American education.*[19]

This combination of input from educators, subject matter specialists, and concerned citizens could be extremely helpful for school districts, both in clarifying and explicitly stating objectives, and in increasing interaction with educational professionals and lay people.

". . . informing the public only of "input information" . . . is not sufficient . . .

Secondly, in the move for accountability, the schools must work toward the formulation of contingencies of reward on the basis of the meeting of responsibilities. It would seem that this is the essence of accountability, but in fact there are no incentives (other than personal) for teachers and administrators in the current system. Reinforcements (salary, other privileges) are delivered to educators at the present time in such a way as experimentation has shown results in low productivity. We would suggest that a better system to use at least in part is modeled on a schedule in which reinforcement is based on productivity, resembling merit pay raises given government workers in some states, and the university system, where salary, tenure and rate of promotion are contingent on productivity. Making rewards contingent on performance has been shown to produce higher rates of response (performance), and in some cases, greater resistance to extinction (perseveration in the face of difficulty, in educational parlance).

". . . a better system . . . is modeled on a schedule (where) reinforcement is based on productivity . . ."

Conclusion

It has been our purpose to take serious note of the move for accountability in American educational evaluation. This movement, in turn, suggests the need for particular data to inform responsible decision-making; the need for assessment instruments which can address the what is learned question; and the need for strengthening the existing ecosystem between our schools, universities, the public and the government agencies concerned with education.

Footnotes

[1] Minneapolis Citizens League, "Accountability in Schools: Not a Threat, but a Real Hope." Minneapolis, November 22, 1972.

[2] S. Jay Samuels, "A Measurement of Learning." The Minneapolis Star, June 27, 1972.

[3] Roger Farr and Nancy L. Roser, "Reading Assessment: A Look at Problems and Issues," *Journal of Reading*, May, 1974, p. 593.

[4] Ralph Tyler, "Why Evaluate Education?" *Compact*, February, 1972, p. 4.

[5] Farr and Roser, *op. cit.*, p. 596.

[6] Benjamin S. Bloom, "Learning for Mastery," *Evaluation Comment*, 1, 2 (May, 1968), p. 3.

[7] *Ibid.*

[8] S. Jay Samuels and Patricia R. Dahl, "Ghetto Children Can Learn to Read—A Personal Report," *The Reading Teacher*, October, 1973, pp. 22-24.

[9] International Reading Association, "Teaching Reading to Special Target Populations," IRA Reports on the Right to Read Effort, 1, 5.

[10] S. Jay Samuels and R. H. Anderson, "Visual Memory, Paired-Associate Learning, and Reading," *Journal of Educational Psychology*, in press.

[11] D. Zeaman and B. J. House, "The Role of Attention in Retardate Discrimination Learning," in N. R. Ellis, ed., *Handbook of Mental Deficiency*, New York: McGraw-Hill, 1963, pp. 159-223.

[12] E.g., Daniel W. Kee and William D. Rohwer, Jr., "Noun Pair Learning in Four Ethnic Groups: Conditions of Presentation and Response, *Journal of Educational Psychology*, in press; William D. Rohwer, Jr., *et al.*, "Population Differences and Learning Proficiency," *Journal of Educational Psychology*, 1971, 62, 1-14; R. B. Green, "SES Differences on Learning and Ability Tests in Black Children," Unpublished master's thesis, University of California, Berkeley, 1969; and I. J. Semler and I. Iscoe, "Comparative and Developmental Study of the Learning Abilities of Negro and White Children under Four Conditions," *Journal of Educational Psychology*, 1963, 54, 38-54.

[13] Frank B. Womer, *What is National Assessment?* National Assessment of Educational Progress, 1970, p. 1.

[14] Carmen J. Finley and Frances S. Berdie, *The National Assessment Approach to Exercise Development*, National Assessment of Educational Progress, 1970, p. 15.

[15] National Assessment of Educational Progress, "States Adopt New Methods," NAEP Newsletter, May-June, 1974, pp. 1-2.

[16] E.g., the Wisconsin Design for instructional use in reading, developed at the University of Wisconsin Research and Development Center for Cognitive Learning, Madison, Wisconsin.

[17] National Assessment of Educational Progress, *op. cit.*, pp. 1-2.

[18] George E. Flanigan, "The Stull Bill—Two Years Later," *UCLA Educator*, XVI, 2(Spring, 1974), pp. 8-9.

[19] Frank B. Womer, *op. cit.*, p. 5.